

Title: What we do with data: a critique of data 'collection'

Author(s): Garfield Benjamin

Abstract (500-700 characters, including blanks): Data collection is everywhere. It happens overtly and behind the scenes. It is a specific moment of legal obligation, the point at which the purpose and conditions of the data are legitimised. But what does the term data collection mean? What does it say or not say? Does it really capture the extraction or imposition taking place? How do terms and practices relate in defining the norms of data in society? This article undertakes a critique of data collection using data feminism and a performative theory of privacy: as a resource, an objective discovery and an assumption. It also discusses alternative terms and the implications of how we describe practices of 'collecting' data.

Short teaser (140 characters): What do we mean when we say data collection? We enact a performative and feminist critique of the term and discuss potential alternatives.

Up to five keywords: Data collection, Feminism, Queer theory, Performativity, Data protection

[Body of the manuscript]

Introduction

'Data collection' plays an important role in representations of data protection and privacy rights. The term is littered throughout privacy policies, consent mechanisms and legislation. It is the point of contact with data subjects at which the generation of data as data (also called datafication - an "obligatory passage point" (Cutting, 2021)) occurs. It is the point at which the context, purpose and consent of data use (processing, transfer, retention, etc.) are formally agreed, with specific status in data protection law (for example, GDPR recital 39 and Art. 13). Historically, the focus on collecting as well as using data had a positive impact, shifting emphasis (especially of rights and regulation) forward from the moment of analysis to the moment at which the data was created. And yet data collection is also only one moment in the middle of a whole series of decisions that determine the power structures under which data is collected. The discourses around data collection have also enacted "political consolidation around the justification for mass data collection [which] marginalized possibilities for fundamental opposition" (Dencik, 2018, p. 37), narrowing the practices available and, through the entrenching of certain practices, cultivating narratives of digital resignation (Draper and Turow, 2019) and disempowerment (Dencik and Cable, 2017) that require shifts in the terms of data collection in order to create spaces for resistance and constructing alternative practices.

Data collection can therefore be a site of oppression and violence, a site in need of refusal and resistance, as well as a site for building communities of and through resistance (D'Ignazio and Klein, 2020; Cifor, Garcia et al., 2019; Benjamin, 2019; Noble, 2018). Patricia Hill Collins (2017) identifies the need to focus on systemic and specific violences even as transversal politics builds intersectional communities of resistance, exemplified in Black Lives Matter founded by three queer Black women, and the inclusion of Black Trans Lives as part of that cause. The violence of data collection is at once administrative and physical, and embedded in everyday life, particularly for people from marginalised groups (Spade, 2015). Community building against such oppression is at once critical and practical, shifting narratives and demonstrating alternative ways of thinking and working (see, for example, Costanza-Chock, 2020).

However, data collection encompasses a wide array of practices, designs and power structures, including issues of regulation and rights, algorithms and automation, imaginaries and implications. The lack of clarity around what data collection is or should be, particularly for the general public, can make rights and obligations vague and opaque, while allowing potentially harmful narratives about the power or value of data to propagate. Against these narratives, it is difficult for individuals and communities, particularly those already marginalised, to examine what data practices are happening to them, or to grasp the full reach of what data collection means. This article therefore enacts a critique of 'collecting data' and its politics, and offers discussion of some alternative terms that may support better data practices. We focus on data collected about people and applied in contexts of social impact, data which largely falls under the legal frameworks of "personal data", while acknowledging that such distinctions are often blurred. The discussion follows a performative model of privacy and data (Benjamin 2020), building on Butler's performativity (1988; 2015) and Sedgwick's periperformative contexts (2003) to consider the normative acts and contextual reception that actively constitute the way data collection is understood and used, towards a more narrative and relational understanding in support of data rights and collective action. It also builds on the seven principles of data feminism (D'Ignazio and Klein, 2020): 1) examine power; 2) challenge power; 3) elevate emotion and embodiment; 4) rethink binaries and hierarchies; 5) embrace pluralism; 6) consider context; and 7) make labour visible. These principles are used to analyse the effects of the terms we use for how data is collected and the importance of considering the inequitable social structures that surround data collection.

This article is structured as follows. Firstly, we discuss the importance of examining the terms of sociotechnical systems and practices, arguing that the narratives that surround such practices are an essential component in understanding data collection. The main body of the discussion is structured around a critique of data collection under three main narratives: data as a resource; data as a discovery;

and data as an assumption. We also discuss the potential for collection as a form of collectives to reclaim power over data. The article then turns to potential alternatives to collection. We focus on creating, curating and compiling data, highlighting the advantages and limitations of each term in relation to collection as well as each other and the broader narratives into which they must fit. The article makes a case for a performative critique of data collection in order to address the constitution of sociotechnical practices through the terms and narratives that surround such practices. We argue that collection is an inadequate, and even harmful term that occludes many of the power dynamics at play. We propose compiling as a more adaptable alternative that balances critique with the redistribution of power. We conclude with further thoughts on the role of terms in constituting sociotechnical systems, and the potential for shifts in terms and narratives to impact on changing sociotechnical norms and practices.

- Terms, practices, narratives

As a term, data collection is linked to the shift in data during the 18th and 19th centuries from natural philosophy and mathematics towards economics and administration (Rosenberg, 2013) or anthropology, all underpinned by the colonial expansion that persists into contemporary data and computational fields (Birhane and Guest, 2020). The specific term "data collection" increasingly saw widespread use across disciplines in the early twentieth century as a method of approaching contemporaneous societal issues including geological surveys for development (Rickert, Hines and McKenzie, 1933, m. 3) or even municipal waste management (Hering and Greeley, 1921). But as a process it is tied to the history of written language as far back as cuneiform and the recording of people and materials for taxation and governance:

Data collection has long been employed as a technique of consolidating knowledge about the people whose data are collected, and therefore consolidating power over their lives. (D'Ignazio and Klein 2020, p. 12)

The shift in meaning of data over time towards something that needs to be "collected", no longer "given" but "found" or "taken", is tied to its more active use in exploiting data to control people and society based on previous observations. This in itself tends towards a preservation of the status quo and the entrenchment of patriarchal, colonial and capitalist aims. It is not that existing regulation is designed to deny individuals their rights, but, rather, that individual rights can do little in the face of the discourses and power structures in which such regulation has been designed. For example, Padden and Öjehag-Pettersson (2021) outline how "the GDPR's framing of 'public interest' privileges economic growth over individual rights" (p. 1), and Tisé (2021) describes "The need to find a "perfect plaintiff" who can prove harm in order to file a suit makes it very difficult to tackle the systemic issues that cause collective data harms", both of which demonstrate the structural barriers to exerting data rights.

Structures such as these enact a wholesale embedding of the desires of the "three Ss: science (universities), surveillance (governments), and selling (corporations)" (D'Ignazio and Klein, 2020, ch. 1). Collection (and its regulation) emphasises the perspective of the collector - and not often critically in mainstream data industry or policy rhetorics. It implies that were data not collected then it would be lost or without purpose, but this carries assumptions that information: a) needs to be operationalised in this way, and b) should be done so according to the purposes decided by already-dominant groups (the three Ss who are most often the "collectors").

The terms we use, and the cultural context of those terms, play a key role in how technologies and practices operate in society:

The metaphors we deploy to make sense of new tools and technologies serve the dual purpose of highlighting the novel by reference to the familiar, while also obscuring or abstracting away from some features of a given technology or practice. (Stark and Hoffmann 2019, p. 5)

The same is true not just for data metaphors, but more generally for the terms used to describe new technologies and practices. This is particularly true for data collection and its integration into a spiral of cultural and political uses and implications that provide a relatable, easily graspable parlance to describe the practice across fields and audiences, but in doing so morphs the meaning and definition of that practice. This is then operationalised to normalise certain definitions and scopes of what data collection is or should be, which in turn shapes regulation and acceptance of often negative data practices. Performativity theory captures this process in which the way individuals speak and act is shaped by cultural norms, but each act also contributes towards constituting those same norms. This can happen inadvertently as certain ways of speaking become entrenched over time. But it can also happen intentionally when more dominant voices with established platforms exert disproportionate influence over the process, for example when business or politics agendas shape marketing and media language that starts to shape the contexts, narratives and expectations of public discourse.

A key contribution of queer performative theory to debates around privacy and data collection is the understanding that terms and practices influence one another. In a more obvious sense, terms can dictate what practices are deemed available or appropriate, while new practices may necessitate new terms. But it goes further than this. Terms and practices cannot be separated. Speech acts are acts, and it is through such acts that meanings are performed. This combines Butler's speech acts and Sedgwick's more literal interpretation of acting and performing. Meanings are spoken and embodied in our practices just as speech is itself a social and political practice. For the present discussion, then, we need terms that allow for critical

description of existing practices but also offer opportunities for constituting more socially just alternative practices.

Applying this thinking to data (and data protection) in particular, the terms that define practices are embodied, entrenched and evolved through those same practices. This establishes and codifies what practices are available or permissible. Intervening in terms can offer one way of intervening with a variety of practices by changing the scope and meaning of those actions. This provides a way for thinking across what Dourish and Anderson (2006) suggested for privacy as practical action and/or discursive practice - under a queer performative framework the two are part of the same social and collective processes, part of the contexts in which data is collected and used. Performativity "describes both the processes of being acted on and the conditions and possibilities for acting" (Butler, 2015, p. 63), and a queer performative theory of privacy thus examines how individual and collective acts are both constituted by and constitute the social norms that define how data and privacy operate in society. Each act of sharing and collecting data (or not) contributes to the expectations around what data should be shared or collected. The periperformative extension to this includes the context in which sharing occurs, as well as a focus on the norms and limits placed around who is able to share or collect data, and who is forced to share, collect or indeed witness data. The queer/feminist analysis presented in this article therefore extends the discussion of concepts such as contextual integrity (Nissenbaum, 2009) and surveillance realism (Dencik and Cable, 2017; Dencik, 2018) by assessing how the terms, norms and practices of data collection are constituted, in order that they might be challenged, changed and reconstituted to address power asymmetries and injustices.

Data collection

The way we talk about data collection constitutes its function in society. The naming of the process plays a performative role in how it is used. Belief in the social narrative of data as something to be "collected" constitutes it as something that can (and even should) in fact be collected. In the discussion that follows, this is examined in how data collection is described, conceptualised and integrated into power structures throughout society.

- Data is not a natural resource

"Data is the new oil of the digital economy" (Toonders, 2014)

"Are you sitting on a data gold mine?" (Khare, 2017)

"Drowning in a data lake?" (Woodie, 2021)

Data is often described as a natural resource to be extracted from the world (Puschmann and Burgess, 2014; Stark and Hoffmann, 2019). The sense of opportunity and risk is externalised into the context of

natural resources, concealing questions of who the data is about, who benefits, and who bears the risks (often unequally distributed). But (especially social) data is not a natural resource, and this metaphor can be both incorrect and harmful. Although this conception of data can highlight the exploitation at work in its extraction from the world, it has come to represent expressions of giddy capitalist colonialism and supports the dominant narratives which enable it to be treated in this way. Alongside a "force" of nature, natural "resources" form one of the major metaphors for data (Puschmann and Burgess, 2014). News reports and tech company marketing is filled with depictions of data as a flood, a lake or a pool. Data as water is "all at once essential, valuable, difficult to control, and ubiquitous" (Puschmann and Burgess, 2014, p. 1699), to which Stark and Hoffmann add the same problem of oil metaphors (another extremely common phrasing that emphasises its value and economic potential) and a history of imperial exploitation of resources designated as "natural". The task of Data Feminism is to examine these unequal power structures and challenge them, including the very terms by which they are used. This is a performative gesture, in the speech acts that designate data as oil or water, and the construction of social narratives that determine how data and privacy are seen and used in different public spheres. For example, it underpins the flawed perception of data as property (Benjamin 2020), which exacerbates power inequalities by assuming data can be traded away.

By speaking data as a natural resource, a performative narrative is developed in which the use of the term collection is attributed specific meanings, tied to processes of extraction. Data collection takes on this meaning as it is perpetuated through its performance across media, research, business and policy contexts. Under this meaning, data is simply out there to be used, and the same colonial practices underpinning exploitation of environmental resources such as oil, water, wood or livestock are justified as part of broader techno-capitalist narratives. But the metaphors are applied in a very one-sided way. Natural resources are emphasised without any inclusion of the stewardship discourses that surround the ecological environment. This "implicitly signals data - and the living people it involves - are open for rank exploitation" (Stark and Hoffmann, 2019, p. 19). But data isn't even treated as carefully as natural resources. For example, Stark and Hoffmann highlight the way data science codes of ethics do not even come close to the detail or scrutiny of the petroleum industry. The ability to copy data endlessly removes the preciousness associated with finding oil or gold. Only data protection rights and social narratives of refusal can provide a level of scarcity for data. Further, data is not uniform: certain types of data, particularly personal data, has more value than others, depending on who the data is about or what it can be used for, and these values feed off and into those rights and social narratives that contribute to the appearance of scarcity and its relation to economic or political value. As Data Feminism implores us to elevate embodiment, we can collapse the metaphor in its material inconsistencies. Data in data collection is

performed as a natural resource that abstracts embodied experiences away from lived realities, hiding the human costs and risks in much the same way that the environmental costs of resource extraction have been hidden through media, marketing and lobbying. But the process of data collection is what constitutes data as data, and therefore as inherently unnatural. We can also make connections with the problematic role of "Human Resources", not only in terms of making labour visible in the collection of data but in data collection as the datafication and operationalisation of individuals, and the gendered, racial and/or ableist biases that perpetuate exploitation of many people and communities by prevailing techno-economic systems.

- Data is not an objective discovery

Contrary to (and even in) the roots of data collection in historical natural sciences, data is not an objective discovery. Feminist, queer, critical race, anti-colonial, anti-capitalist and intersectional approaches push against the myth of completeness that surrounds the scientific narratives of data. Such narratives only serve to objectify people - as individuals, communities and populations - and add them to the collections of the powerful. Data is not simply discovered in the world and collected for use. This conceptualisation fails to afford adequate representation to the labour involved. The use of collection as a term can remain useful to talk about labour and its erasure in the historical narratives of computing (Hicks, 2018) by suggesting that it is companies collecting from their workers' labour rather than directly from data subjects. However, this labour is not merely collection but production: data "collection", by implying discovery, fails to capture the productive qualities of data work in the "data supply chain" (Posner, 2018). This formulation of data as discoverable and collectable is underpinned by a false sense of objectivity, which supports historical and existing technological, colonial (Appadurai, 1993), patriarchal and capitalist power structures that define the data and computational sciences (Birhane and Guest 2020).

As D'Ignazio and Klein's Data Feminism asserts, "data are not neutral or objective. They are the products of unequal social relations, and this context is essential for conducting accurate, ethical analysis" (2020, ch. 6). Instead, we should embrace the incomplete and the incomputable (Hildebrandt, 2019). In the messiness of data lies the importance of incorporating "data settings" (Loukissas, 2019), emphasising localities and considering the contexts in which data has been collected (Noble, 2018, 149; D'Ignazio and Klein, 2020, ch. 5). For example: the performative power of gendered language can be harmful and exclusionary in health settings, making it difficult for certain individuals to get the help they need, particularly around binaries but also looking beyond categories to spectra of gender and sexual identities (Bouman et al., 2017); beyond the one-size-fits-all attempts to "fix" processes of collecting data about protected characteristics, the nuances of different categories of racialisation might be appropriate for different settings or purposes (Hanna et al., 2020);

and issues of when to include different aspects of data in legal claims underpins the argument for intersectionality (Crenshaw, 1989). Building on these engaged and contextually-aware social science approaches, there is a critical periperformative gesture - a narrative about the narrative, a (re)setting of the terms under which critique can be performed - that is needed to constitute data as always being embedded within specific contexts, always embodying specific human lives.

Hoffmann calls us to "resist uncomplicated claims to neutrality or objectivity" (2018, p. 7), and this issue is particularly prominent in the datafication of identity. Queer theory acts as an important tool in its "radical anti-identity politics [that] rejects a stable, knowable subject" (Green, 2007, p. 29). Similarly, in intersectional Black Feminist information studies, "control over identity is political and often a matter of public policy" (Noble, 2018, p. 135). On the one hand, data collection can perform and thereby impose fixed categories (such as gender and/or sexual identities and/or expressions) on people for whom such categories do not apply, while on the other hand, lack of acknowledgement of intersecting categories (such as being both Black and a woman) can perform and thereby impose deficit narratives. In both cases, people's identities and experiences are often erased. These perspectives demand a critique of oppressive tools such as "menu-driven identities" (Nakamura, 2002, p. 101-2) that produce and reproduce binaries and hierarchies. When we understand datafication of gender labels, for example, as contextual and shifting over time (Szulc, 2020), or racial categories as colonial and technological constructs (Hanna et al., 2020; Birhane and Guest, 2020; Benjamin, 2019), then simply adding more options to the menu, or even free text boxes, is neither subversive nor adequate. It is the act of data collection conceived as a specific gesture performed at a specific point in time that is carried forward into (the illusion of) an objective fact that reproduces inequalities.

Existing precarities (including of identity such as gender, race or class) are closely linked to levels of identification over identity (Wood, 2017, p. 45), in which agency over one's own life and experiences is subsumed under the imposition of categories for administration, monitoring or control by dominant actors. But equitable representation is nothing without systemic transformation of the underlying structures and narratives (D'Ignazio and Klein, 2020, p. 31). Data feminism enacts a refusal of perpetuity based on one-time consent, of shallowly disruptive "venture discourse", and of harmful data regimes and logics that valorise the objective as a thin moral cover (Cifor, Garcia et al., 2019). As Ruha Benjamin writes:

The employment of new technologies that reflect and reproduce existing inequalities but that are promoted and perceived as more objective or progressive than the discriminatory systems of a previous era. (Benjamin, 2019, p. 5-6)

Examples abound, particularly in the collection of data for use in algorithmic decision-making. For example, the 2020 UK exam results debacle demonstrates an algorithm working as intended, but based on the assumption that past geographical data was a valid method of assigning current grades by restricting students to the performance of previous attendees of their school, with protests and U-turns demonstrating failures of standardisation (and the use of data to achieve this) and of policy-making (and the decision to apply past data in the first place) (Kippin & Cairney, 2021). This is a performative process in which dominant oppressive narratives constitute the belief in data collection as discovering objective truths. Datafication - which is the transformational process of data collection on individuals, groups or society - is constituted on the basis of "problematic ontological and epistemological claims", existing "between scientific paradigm and ideology" (van Dijk, 2014, p. 197), and itself constitutes social reality according to the logics of the data used. Beneath debates over subjectivity and objectivity lies the normative dimension inherent to the design of databases and the decision to collect data in any specific way (Benjamin, 2019, p. 78). The performance of empiricism, by making shallow gestures towards inclusion or fairness, therefore constitutes objectivity narratives in which inequitable and unjust practices are deemed acceptable by imposing certain expectations around what data collection is and does, and thereby control assumptions of when and how data collection should be done.

- Data collection should not be assumed

If the data that is collected is not neutral, neither is the process of collecting data. It should never be uncritically assumed that data collection is an appropriate course of action. Feminist data protection enacts a key moment of refusal in data collection (Cifor, Garcia et al., 2019). The act of collecting data is too often an act of violence that needs resisting. Data collection is epistemic violence (Spivak, 1988, p. 24), administrative violence (Spade, 2015) and data violence (Hoffmann, 2020); it defines, reduces and restricts personhood, denying certain groups (such as women and/or Black, Indigenous, queer or trans people) the agency to exist as a person with lived experience embedded within relational contexts. "Collection" doesn't capture this erasure, the removal of the person. This is among the reasons that marginalised communities (Lewis, 2017; Petty et al., 2018) have built conceptions of privacy on the disability activism assertion "nothing about us without us".

However, the term data "extraction" also fails to capture the full extent of this violence. It is not merely something being taken out of an individual, it is a complete refusal of a person as a person, it is data as exclusion. But there is also more to this violence than erasure. Reducing people to categories - the datafication and objectification inherent to data collection as conceived by "science, surveillance and selling", that is, academic, government and corporate interests (D'Ignazio and Klein 2020, 42) - imposes something external

in its drive towards identification. Datafied identities are not only performed according to cultural norms, binaries and hierarchies, but explicitly scripted to fit machine-readable periperformative contexts built on patriarchal, racist and ableist inequities.

In this sense, "inclusion" can also be an act of discursive violence (Hoffmann 2020), unless it is joined by challenges to oppressive power structures and hierarchies. An obvious example is when selecting categories from a limited menu is required to access a service. The price of entry can impose, for example, a binary choice of gender identity that fails to take into account the wide spectrum of identities that exist. But more widely, recognition of more options for gender identity - while seemingly inclusive - places a cost of fixing one's identity at a certain point in time, the moment when that data is collected, even though identities may be fluid over time (see, for example, Ruberg and Ruelos, 2018) while the fixed data persists in representation and decision-making. Similarly, if facial recognition systems that fail to recognise darker skin tones are improved, this could lead to increased targeting of racialised groups when such technologies are applied in social contexts with existing issues of discrimination such as policing. This can be applied to the setting of data collection but also the process itself: inclusion within a dataset can be an extremely violent and oppressive act, whether that is an administrative attack on one's identity or a physical attack either when that data is used or when it is collected in the case of, for example, discriminatory arrest practices.

This aligns with Andrejevic's conception of data surveillance as the creation of digital enclosures (2009), and datafication therefore as imprisonment within the societally performed roles of, for example, fixed gender or racial identities. The process of identification that is assumed in data collection is itself a technological tool for power (Benjamin, 2019). We need sociotechnical tools to examine and challenge the normalisation of these racist, sexist, ableist, homophobic and transphobic narratives by disrupting the assumption of data collection as necessary. The systemic violence of data collection (as extraction, exclusion and imposition) is built into online technical, legal, political, economic and social architectures. We must bear in mind that many social relations are not voluntary (Losh, 2015, p. 1651; Hoffmann, 2018, p. 11), with users often being condemned to engage (Rider and Wood, 2018). For example, Facebook tries to constitute a periperformative context in which users have "no expectation of privacy", and uses this to misdirect regulation and resist legal accountability in court, such as against a class action brought in relation to the Cambridge Analytica scandal (Thalen, 2019). The asymmetric agency in the performance of data collection creates oppressive assumptions about what data can do, when it should be used, whose interests it should serve, and, underpinning these, who decides what data collection means in society. Against these narratives, alternative terms and practices are required that see "contemporary (often undesirable) mass data collection as a contingency that has been

actively constructed as an inevitability, which can therefore also be challenged and reconstructed" (Dencik, 2018, p. 41). Feminist data protection must examine and challenge the terms, practices and assumptions at work, and generate new narratives and contexts that redistribute the terms under which data is collected.

- Data collectives

One useful way the term data collection can be reclaimed when talking about communities and data is to speak instead about data collectives, and collective aspects of data collection. There is the possibility of a radical performative assembly (Butler 2015) through collecting data together. This leans on the empowering "we" behind individual acts, an assertion of collective agency that might use data in ways that emphasises the Data Feminist principles of elevating emotion, embracing context, promoting pluralism and acknowledging labour. Not only is data collection usually defined in terms of data protections of individuals, rooted in digital economics, privatisation and retributive justice rather than social justice and the redistribution of power and wealth, but these legal frameworks also tend to treat individuals in the abstract (Strycharz, Ausloos and Helberger, 2020). We therefore see norms established that fail to appreciate the interwoven nature of who data is about. Data is a collective performance. Data terms and practices are constituted socially not individually. Data is relational, and data collection and protection also need to be conceptualised in a relational way.

Social data in particular is always about multiple people (Sarigol, Garcia and Schweitzer, 2014), even when legally it is usually considered as belonging to one individual (identified also in technological responses and the closing off of resistance around mitigating harms rather than more radical social change - see Dencik, 2018, p. 39). For example, even if data is only about one person, such as facial biometrics data, this can still have collective implications. One individual may not suffer discrimination through the use of an image of their face collected to train a facial recognition system, nor even suffer the effects of its unequal deployment by law enforcement (often slipping through loopholes in regulation or even falling under a perverse performance of data altruism). But the fact that this particular face has been collected in a way that constitutes a "false norm" (Giesecking, 2018, p. 150) against which already marginalised groups are further discriminated renders it an issue of collective privacy, collective identity and a relation with the specific context or environment that goes beyond conceptions of individual privacy rights (Käll, 2019). This returns to the principles of examining and challenging power, and we need to bear in mind that while existing data protections focus on individuals, data is often collected by large organisations. As Mantelero warns us, in data protection "collective interests require adequate forms of representation, as well as the involvement of a range of stakeholders in the balancing of conflicting interests" (2016, p. 254). And a limitation of collectives that we must

consider is that they are not single entities but are themselves contexts with internal differences, social relations and potentially asymmetric structures.

Thinking about collective data ethics and data protections must also lead towards more representative governance of data throughout the ecosystem and lifecycle. Existing collective mechanisms, such as collective redress (including via non-profits) through complaints to Data Protection Authorities as afforded by legislation such as the GDPR (article 80), remain reactive and unsatisfactory, doing little to prevent harms nor to redistribute decision-making surrounding the collection of data. The EU's proposed Data Governance Act makes some progress towards more collective approaches, placing greater emphasis on the roles of data intermediaries and enforcing a duty of care on such organisations. Directive 2020/1828 provides for the somewhat more proactive injunctions, but is still reactive to harms that have already occurred (even if they have not happened to a given specific individual), rather than, for example, the FDA-for-algorithms approaches (Tutt, 2017) which would require much more significant design- and deployment-side regulatory mechanisms. Even these measures could go further in a "relational turn" to do more to address power asymmetries and collective interests (Richards and Hartzog, 2020, p. 5). Otherwise they risk continuing to emphasise the relationship between the individual and the decision-maker (now shifted to data intermediaries), even if that is to increase agency it is personal agency. Instead, emphasis should be placed on the broader social concerns with what, how and why data is or should be collected. But data intermediaries can also be seen as rights-management-as-a-service (converting human rights into the frameworks of copyright for which regulators such as the EU demonstrate a continued favour). It is disheartening to see the further embedding of business interests and power asymmetries in the EU's proposed Data Governance Act, in which data sharing service providers (data intermediaries) receive greater powers to trade data on subjects' behalf and thereby transfer rights to such intermediaries (EDPB, 2021), whereas subjects are explicitly restricted from transferring rights to data cooperatives (which arguably embody more collective interests and decision-making). However, caution of transferring rights - rather than more general decision-making - remains an important risk to take into account. Data intermediaries also enact a shift in norms which further the imposed expectation of sharing (often leaning on, for example, concepts such as data altruism in the proposed DGA, or the lack of systemic change to exploitative models in that or the GDPR). These measures to entrench fundamentally conservative approaches to data protection enact a periperformative shaping of the context that places increased burden on people(s) to share rather than on organisations to justify the need for data. This boils down to political issues - as is "every single decision and action around data" (Bartoletti, 2020) - and therefore cannot be fixed by law alone - hence our focus in this article on social norms and practices as sociopolitical approaches.

Adopting more feminist practices, "when we count within our own communities, with consideration and care, we can work to rebalance unequal distributions of power" (D'Ignazio and Klein, 2020, p. 123). A consideration of care ethics is useful. A duty of care goes beyond accountability, as it requires more proactive action in the interests of data subjects, and care itself goes beyond merely the duties of data controllers/processors to data subjects. Dean Spade (2020), as well as Hil Malatino (2020) building on Leah Lakshmi Piepzna-Samarasinha (2018), root trans, feminist, queer and crip theories of care in mutual aid. Shifting towards mutual ownership, mutual decision-making and need enacts a significant challenge to existing power structures, and an opportunity to constitute new contexts within which data operates and decisions about data are made. It is about building new systems towards collective power and justice. Care ethics in data creates a collective space to acknowledge particularity and empathy (Fotopoulou, 2019), and situate knowledge in feminist epistemic frameworks (Luka and Millette, 2018). It asks the feminist questions of who data is collected about, for and by. D'Ignazio and Klein (2020, ch. 5) highlight within this data feminism a commitment to a multiplicity of voices - building on Haraway's conception of knowledge as partial and situated knowledge (1988) - as the path towards a more complete (though never to be considered whole) picture of an issue in all its social complexity. The critical process can thereby lead towards "optimistic gestures that can imagine alternative sets of social relations" (Luka and Millette, 2018, p. 5). By "working with data subjects rather than capturing data objects" (Cifor, Garcia et al. 2019), acts of refusal become acts of commitment to using data to perform and thereby constitute more equitable social structures. Ways forward include developing expectations and practices of knowledge transfer in both directions as well as building social infrastructure (D'Ignazio and Klein, 2020, ch. 5), in which we "imagine our end point not as "fairness", but as co-liberation" (p. 53). Data collectives are local and global, they are relational, plural, contextual, messy, empathetic, critical and optimistic frameworks within which intersectional feminist principles can lead towards co-liberation of, from and through data. Thinking in this way moves into broader and more social conceptions of performing privacy (Benjamin, 2020) that cannot be captured within the legislative frameworks of data protection alone.

Alternatives to data collection

In the preceding section, we have examined existing meanings constituted by the term data collection. We discussed narratives of data as a natural resource (legitimising exploitative extraction practices), objective discovery (concealing oppressive motives and practices under the guise of neutrality), an assumed part of contemporary society (normalising surveillant power asymmetries), and data collectives (including the individualising and disempowering norms of existing regulation and discourse). If data collection remains useful as a term only when talking about the possibilities for collective action that might reframe the power structures surrounding

data, what term can we use for the act, the gesture, the performance of collecting data? The terms we use are tied in with legitimising specific sets of practices, and creating wider expectations about what data is and who it is for. We turn the discussion now towards possible alternatives to the idea of "data collection", to ask what other priorities, principles and practices we can perform through the terms we use, as well as how they might enable shifts in the contexts and narratives surrounding data.

- Creating data

One approach is to talk about creating data, highlighting the production of data as an active and labour-intensive process. This emphasises the final principle of Data Feminism in making labour visible, particularly when workers are minoritised, disempowered or subjected to highly asymmetric power structures. Constructing data may also be considered along similar lines, emphasising the politicisation of the labelling as evidence or objective facts. Creating data also emphasises the narrative or interpretative processes that actively separate data from embedded contexts and lived experiences. Data might be prepared in order to be read by machines, but it remains a process constituted by human decisions. And as the Feminist Data ManifestNO asserts, data must be "acknowledged as at once an interpretation and in need of interpretation" (Cifor, Garcia et al., 2019). Constructed data asks for post structural critique to confront the binaries and hierarchies that surround its construction and use(s). It is not enough to say that data collection is an interpretative act, as that risks giving too much weight to the falsely objective processes of science, governments and big tech. Data that is created must be performed, and it must have a periperformative audience (Sedgwick, 2003). When used critically, this audience exists not in order to be "forced to bear witness", and accept claims of objectivity, but to support the collective and affective qualities of data, to emphasise context. Creating data also adds potential issues in a legal sense, as it risks inferring ownership of data as intellectual property under the control of corporations - the "settler colonial logics of data ownership" (Cifor, Garcia et al., 2019) - raising issues of attribution, decontextualisation, fictionalisation and the erasure of relationality, sensitivity, context, affect or stewardship. Creating data allows space for co-creation and co-liberation under more equitable power relations and just design practices. But it also risks underplaying or blurring the different power structures, roles and responsibilities within the co-creation of data between data subjects, human labour, platforms and social narratives.

- Curating data

If creating or constructing data does not encompass the different processes at work, curating data offers an expanded alternative. We are considering not just an active, creative, performative gesture that brings data into being as data, but also integrating the interpretative

gesture as a process of framing or editing. This emphasises the reductive or abstractive processes that go into data collection. There is an element of consolidation always implicit in this process, of bringing together and comparing multiple people to make overarching assumptions and generalisable comments. We should be wary of thinking too far in terms of consolidation, as it risks the human resources approach that objectifies people to their function within a dataset. Instead, curation leans on the creative processes involved, but in both senses we see an act of reframing that also always translates lived experience into machine-readable formats.

Curating data also emphasises the distribution of labour: collecting data, the task of curation, has its own work involved, its own labour force and exploitation; but this does not diminish the creative, emotional and affective labour of data subjects that is converted into data. The "artist" or "performer" remains the data subject, while the "curator" is those collecting the data. Similarly, this conception leaning on creative economies gives space for "gallery directors" or "stage managers" - the platforms or governments with disproportionate power to shape the periperformative contexts in which the creators and curators are expected to work. Even the data intermediaries hailed by the proposed Data Governance Act risk occupying a similar position of influence over norms and expectations, as discussed above. These unequal power relations make clear the restrictions that are applied in data collection, the narrowing of scope to fit economic interests. They also show the broader narratives that define the assumptions of the audience of data, and the different types of roles that we are expected to perform.

Useful in curation, referring to both the potential for data as a natural resource in need of protection and in the onwards responsibility of curators towards works of art, is the concept of stewardship. Data stewardship has been conceived as "environmental data curation" (Baker and Yarmey, 2009). This emphasises responsibility and care within a curatorial context. It also highlights the dangers of viewing data as a natural resource, and asks us to embed data within a lifecycle and ecosystem that requires relational care, which, as discussed above, can lead also towards redistributive economic and power structures. Stewardship suggests onwards data protections, including of groups, which can support how particularly marginalised communities are treated using data as well as enforcing time-based protections such as the right to access and the right to be forgotten. These are important aspects of how data is kept, shared and used. While stewardship is not specific to curation, a curatorial conceptualisation of data collection establishes a narrative in which those who collect data take responsibility to care for it. However, as we have seen in the different roles implied by curation, it is not necessarily those creating the data - often built off low-paid and exploited labour - who necessarily hold power. Curation, like creation, risks perpetuating "Eurocentric canons" embedded with "racism, patriarchy, capitalism, homophobia, transphobia, and ableism" that still plague the neoliberal

art world (Balona de Oliviera, 2020, p. 80). It is those stage managers, those gallery owners - the platforms, corporations and governments who seek to hoard ever more data and use it to their own ends - who must be held accountable. Justice combines with care here to create narratives of responsibility, but this can never be left into the hands of, for example, corporations to self-regulate. Local, national and international regulation of curation and stewardship is required, with community involvement centring those most affected. If a curatorial model is to be applied, it must be that of a community-led and co-creative arts practice that redistributes power to the "artists", or data subjects, who are most at risk. Curation is a useful description, particularly within a performative framework already leaning on the wider contexts of creative practices, but it is still not quite sufficient.

- Compiling data

Instead, we suggest compiling as a useful alternative. This has the advantage of leaning on an established computing term with connotations of translation from higher to lower order languages, emphasising the reductive process of generating data from the world and converting it into a machine-readable format. But primarily it can be understood in terms of an elaboration of curatorial conceptions (as in compiling an anthology) that more closely embodies data feminist principles such as embracing multiplicity by addressing some of the limitations of curation around labour and power. The emphasis of compilation on a more explicit relation with economic and editorial aspects embodies more clearly not just a performative dimension of data but also the periperformative dimensions of setting contexts and norms. This also helps keep a critical eye on the problematic histories of creation and curation, while still emphasising the non-objective nature of data compared to collection. For example, the role of the collector (itself readily evoking colonial practices of forcibly acquiring cultural artefacts to place in Western museums) may allow for the subjectivity of choosing what to collect, but compiling provides an extra level of engaging with what is chosen for both inclusion and (perhaps more importantly) exclusion in the always-already-edited practice of defining the (power) structures of data sets. Compiling integrates elements from collection, creation and curation, but offers a more open term that allows for the performing of alternative meanings and practices.

The historical impact of data compiling has been seen as setting accepted standards (Rossini, 1967), a performative gesture, as well as mathematising the dynamics of the social (Barnes and Wilson, 2014), and more recently in a sense of taking stock of complex issues and crises (Bedford et al., 2018). This suggests a flexible set of meanings that already engage with performative aspects of how terms are used. Today, popular, technical or business definitions emphasise the mediating process of compilation - and making data for a purpose. This ties in with the expectation of specific uses for data and its interpretation,

perhaps offering a way of reducing the drive towards data collection for its own sake that dominates current practices of platforms and data brokers. Sheila Jasanoff (2017) associates compilation with problematisation, using the term for singular or plural processes and somewhat interchangeably with collection. But her discussion pushes towards compilation as associated with assemblage, which helps us emphasise the social origins, impact and relations of data practices. Implicit in Jasanoff is a use of compilation as more systematic than collection and, while taking a critical perspective of the "political choices that accompany any compilation of authoritative information" (p. 12), she describes how "the story of data compilation can also be told in positive terms as one of growth in the capacity of human societies to generate systematic knowledge about themselves" (p. 8).

Taking a step back, the more general, metaphorical definitions of compilation are also multiple:

- 1: to compose out of materials from other documents
\\compile a statistical chart
- 2: to collect and edit into a volume
\\compile a book of poems
- 3: to build up gradually
\\compiled a record of four wins and two losses
- 4: to run (something, such as a program) through a compiler
(Merriam Webster, 2020)

Immediately, we see statistical and temporal uses. Compilation occurs when bringing together other sources, such as lived experiences from the world but also potentially different types or sources of data. It also happens over time, emphasising the need for spatiotemporal considerations not only of the context in which data is compiled but also the continued stewardship and data protections required afterwards. There are potential limitations to the term compilation, such as a risk of de-emphasising the extractive processes involved, but the etymological root of compiling in the Latin "to plunder" highlights the exploitative potential of the process.

Maintaining the implied creative processing from curatorial perspectives, compilation leans on creative metaphors, as in a compilation of songs, stories or poems. It shows that datasets are a grouping together of individual people, individual stories, individual experiences, but that they are also always changed in their new framing as part of the dataset. For example, a song removed from a concept album might lose much of its meaning, just as data without context removes a significant amount of social complexity. Equally, a song may gain something when added to a playlist for a specific purpose, just as data can be brought together to build communities and create the basis for action. The "plundering" etymological origins of compilation also highlight the potential for creative exploitation when a song is reframed and recommodified in, for example, a "best of" compilation. Where this might assign a certain age, genre or other label, the

process reflects the imposition of identification that harms a person's ability to perform their identity over time in the context of social relations rather than datafied categories.

Using compilation seeks to avoid blurring the facets of creative control and the assumption of consent, involvement (including labour and decision-making) or benefit (and, conversely, harms). Compiling data is an act of editing and translating narratives of lived experience; it acknowledges co-creation between subject and collector while highlighting differences in roles, labour and audiences as well as power asymmetries. Where curation leads into the appearance of a positive creative process, revealing a tension between neutrality and authorship underpinned by historical sexist and racist discriminations, compilation also highlights the potential commercial interests that impose on the creative gestures and wrap them in dominant narratives. In a positive sense compiling can come to emphasise caretaking or stewardship of data and the performative potential of re-presenting data, particularly in the periperformative processes of shifting contexts.

While compilation embraces elements of curation, and a creative process of putting data sets together, it also brings a technical meaning. In computing, compiling is something that happens to code to make it work, converting it from the human coding language into something a machine understands and can make happen. The link to assembly highlights the politics of compiling and relational assemblages of labour and data ecosystems, emphasising the breadth and types of information collected about individuals across platforms as well as their inclusion in mass data sets. Like data collection, compilers enact a translation from human narratives into machine readable, countable pieces of information that can be searched or processed. Compiling is also a time when programmes can go wrong - the compiling fails because of errors not in the data - as defenses of biased AI systems such as face or emotion recognition would have us believe - but errors in the code and in the narratives of code. We can refer back to Ruha Benjamin's emphasis on codes as narratives, "telling us what to expect" and operating "within powerful systems of meaning that render some things visible, others invisible, and create a vast array of distortions and dangers" (2019, p. 6-7), in other words as social technologies. This is important to remember in the flawed transition from social realities to data. Leaning on this trope of compiler errors, we must assert within the use of any term the non-neutrality of the process and promote the social over the technical. Compilation still requires conceptual work to remain useful across contexts, but it usefully brings together positive and negative, creative and technical, labour, relational and contextual framings of the processes usually called data collection.

Conclusion

This article has presented a critique of the terms of data collection and the norms and expectations such terms engender. Through the

contribution of queer intersectional feminist theories, leaning on the concepts of performativity and periperformativity, we have issued a challenge to the way data terms and data practices continually constitute one another. Challenging the terms used raises fundamental questions about what data collection is or can be, including who it is about, by and for. In particular, it asks whether data collection should happen at all. Can data collection be positive? Should it be protected? What are the contextual specificities? In the state of its current dominant narratives, data collection is performed as an exploitative and oppressive act. But it can be conceived, constituted and performed otherwise.

D'Ignazio and Klein (2020) outline a range of historical cases and contemporary community-driven projects that embody feminist data principles - whether it is revealing gender promotion barriers at NASA, demonstrating the racialised distribution of child fatalities on school routes, reclaiming maps with indigenous place names in Canada, or building an Atlas of Caregiving to make visible the effort and expertise of carers - where data has been used effectively to challenge power and hierarchies, and to enable change by promoting context, emotion and labour. Critical race and intersectional approaches by Benjamin (2019), Noble (2018) and Costanza-Chock (2020) further demonstrate the conceptual sociotechnical tools required for these processes.

A performative critique of data collection, and a reframing under alternative terms such as data compilation, highlights how there is always a narrative component to data. But these narratives constitute dominant - oppressive, patriarchal, racist, colonial, ableist, homophobic, transphobic - values into the intertwined performance of meanings and practices, which in turn defines the debates, expectations and regulation of such terms and practices. With each gesture that performs and reinforces these narratives, exploitative and asymmetric power structures are further legitimised. Important in a performative critique is developing new gestures that shift the context of discussion and create radical contexts in which alternative narratives can be developed to deconstruct existing binaries, hierarchies and power structures. The choice of term, therefore, should allow space for critically inspecting the practices and processes by which data is produced, but emphasise a shift towards practices that are equitable and just.

Proposing terminological shifts can impact upon how we conceive data rights, data justice and data activism, as well as considerations for the language of future regulation to emphasise such shifts more in favour of data subjects than platforms, not just in terms of individual protections or even collective redress, but by redistributing decision-making and benefits, through potential methods of collective performative action. Speech acts - and the terms we use to enact them - constitute practices and sociotechnical imaginaries, shaping the possibilities for redistributing power and justice. The way we talk

about issues is intertwined with our expectations and practices, and the way these issues are performed in society. In this article we have offered a critique of data collection through performative analysis and the principles of data feminism. The proposed replacement with "data compilation" provides a way of examining power in the different roles and priorities of who is involved in the different levels of compilation. It thereby offers a way to challenge power by holding responsible the compilers for translations, errors and ongoing stewardship. It elevates emotion through emphasising creative processes and the active role of data subjects. It addresses binaries and hierarchies in highlighting the reframing and imposition of certain categories. It embraces pluralism by focusing on the bringing together of different data subjects into new collectives. It invites us to consider context in the framing of a compilation against the original creative situation. And finally, it makes emotional and technical labour visible in relation to power structures, responsibility and agency, in order that collectives and communities might repurpose data compilation to create alternative narratives of data compilation towards more equitable and just purposes.

[Bibliography]

Andrejevic, M. (2009). Privacy, Exploitation, and the Digital Enclosure. *Amsterdam L.F.*, 1, 47-62.

Appadurai, A. (1993). Number in the postcolonial imagination. In C. Breckenridge & P. Van der Veer (Eds.), *Orientalism and the postcolonial predicament* (pp. 314-339). University of Pennsylvania Press.

Balona de Oliveira, A. (2020). Breaking Canons: Intersectional Feminism and Anti-Racism in the Work of Black Women Artists. *Vista*, (6), 79-100.

Baker, K. S. and Yarmey, L. (2009). Data stewardship: Environmental data curation and a web of repositories. *Digital Discourse: The Evolution of Scholarly Communication*, 1(1), 12-26.

Barnes, T. J., & Wilson, M. W. (2014). Big data, social physics, and spatial analysis: The early years. *Big Data & Society*, 1(1), 1-14.

Bartoletti, I. (2020). *An Artificial Revolution: On Power, Politics and AI*. Indigo Press.

Bedford, J., Gercama, I., & Bardosh, K. (2018). Social Science and Behavioural Data Compilation–November 2018. Institute of Development Studies. <https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/14144>.

Benjamin, G. (2020). From protecting to performing privacy. *The Journal of Sociotechnical Critique*, 1(1), 1-30.

- Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity.
- Birhane, A., & Guest, O. (2020). Towards decolonising computational sciences. *Danish Journal of Women, Gender and Research*.
- Bouman, W.P., Schwend, A.S., Motmans, J., Smiley, A., Safer, J.D., Deutsch, M.B., Adams, N.J. & Winter, S. (2017). Language and trans health. *International Journal of Transgenderism*, 18(1), 1-6.
- Butler, J. (1988). Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory. *Theatre Journal*, 40(4), 519-531.
- Butler, J. (2015). *Notes Toward a Performative Theory of Assembly*. Harvard University Press.
- Cifor, M., Garcia, P., Cowan, T.L., Rault, J., Sutherland, T., Chan, A., Rode, J., Hoffmann, A.L., Salehi, N., Nakamura, L. (2019). Feminist Data Manifest-No. <https://www.manifestno.com/>.
- Collins, P.H. (2017). On violence, intersectionality and transversal politics. *Ethnic and Racial Studies*, 40(9), 1460-1473.
- Costanza-Chock, S. (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need*. MIT Press.
- Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum*, 1989(1), 139-167.
- Cutting, K. (2021). "Let us use our power": moving beyond datafication in technology design. Data Justice 2021, Cardiff University, 21 May 2021.
- Dencik, L. (2018). Surveillance realism and the politics of imagination: is there no alternative?. *Krisis: Journal for Contemporary Philosophy*, 2018(1), 31-43.
- Dencik, L., & Cable, J. (2017). The advent of surveillance realism: Public opinion and activist responses to the Snowden leaks. *International Journal of Communication*, 11, 763-781.
- D'Ignazio, C. & Klein, L. (2020). *Data Feminism*. MIT Press.
- van Dijk, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197-208.

Dourish, P. & Anderson, K. (2006). Collective Information Practice: Exploring Privacy and Security as Social and Cultural Phenomena. *Human-Computer Interactions*, 23(3), 319-342.

Draper, N.A., & Turow, J. (2019). The corporate cultivation of digital resignation. *New Media & Society*, 21(8), 1824-1839.

EDPB (2021). Statement 05/2021 on the Data Governance Act in light of the legislative developments. European Data Protection Board.
https://edpb.europa.eu/system/files/2021-05/edpb_statementondga_19052021_en_0.pdf.

Fotopoulou, A. (2019). Understanding citizen data practices from a feminist perspective: embodiment and the ethics of care. In H. Stephansen & E. Trere (Eds.), *Citizen Media and Practice*. Taylor & Francis/Routledge.

Giesecking, J.J. (2018). Size matters to lesbians, too: Queer feminist interventions into the scale of big data. *The Professional Geographer*, 70(1), 150-156.

Green, A. (2007). Queer Theory and Sociology: Locating the Subject and the Self in Sexuality Studies. *Sociological Theory*, 25(1), 26-45.

Hanna, A., Denton, E., Smart, A., & Smith-Loud, J. (2020). Towards a critical race methodology in algorithmic fairness. *ACM FAT* 20*, 501-512.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist studies*, 14(3), 575-599.

Hering, R., & Greeley, S. A. (1921). *Collection and disposal of municipal refuse*. McGraw-Hill.

Hicks, M. (2017). *Programmed inequality: How Britain discarded women technologists and lost its edge in computing*. MIT Press.

Hildebrandt, M. (2019). Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law*, 20(1), 83-121.

Hoffmann, A.L. (2018). Data, Technology and Gender: Thinking About (and From) Trans Lives. In J.C. Pitt and A. Shew. *Spaces for the Future: A Companion to Philosophy of Technology* (pp. 3-13). Routledge.

Hoffmann, A. L. (2020). Terms of inclusion: Data, discourse, violence. *New Media & Society*.

Jasanoff, S. (2017). Virtual, visible, and actionable: Data assemblages and the sightlines of justice. *Big Data & Society*, 4(2), 1-15.

Jannice Käll (2019) A Posthuman Data Subject? The Right to Be Forgotten and Beyond, *German Law Journal*, 18(5), 1146-1162.

Khare, R. (2017). Are you sitting on a data gold mine? What all brands need to know. Global Marketing Alliance. <https://www.the-gma.com/data-gold-mine>.

Kippin, S. & Cairney, P. (2021). The COVID-19 exams fiasco across the UK: four nations and two windows of opportunity. *British Politics*, 1-23.

Lewis, S.J. (2017). *Queer Privacy: Essays from the Margins of Society*. Mascherari.

Losh, E. (2015). Feminism Reads Big Data: "Social Physics," Atomism, and Selfiecity. *International Journal of Communications*, 9, 1647-1659.

Loukissas, Yanni A. (2019). *All data are local: Thinking critically in a data-driven society*. MIT Press.

Luka, M. E., & Millette, M. (2018). (Re) framing big data: Activating situated knowledges and a feminist ethics of care in social media research. *Social Media + Society*, 4(2), 1-10.

Malatino, H. (2020). *Trans Care*. University of Minnesota Press.

Mantelero, A. (2016). Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection. *Computer law & security review*, 32(2), 238-255.

Merriam Webster (2020). Compile. <https://www.merriam-webster.com/dictionary/compile>.

Nakamura, L. (2002). *Cybertypes: Race, ethnicity and identity on the Internet*. Routledge.

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.

Noble, S. (2018). *Algorithms of Oppression*. NYU Press.

Padden, M. & Öjehag-Pettersson, A. (2021). Protected how? Problem representations of risk in the General Data Protection Regulation (GDPR), *Critical Policy Studies*, 1-19.

Petty, T., Saba, M., Lewis, T., Gangadharan, S.P. & Eubanks, V. (2018). Reclaiming our Data: Interim Report. Our Data Bodies. <https://www.odbproject.org/wp-content/uploads/2016/12/ODB.InterimReport.FINAL.7.16.2018.pdf>.

- Piepzna-Samarasinha, L.L. (2018). *Care Work: Dreaming Disability Justice*. Arsenal Pulp Press.
- Posner, M. (2018). See No Evil. *Logic Magazine*, 4.
- Puschmann, C. & Burgess, J. (2014), Big Data, Big Questions| Metaphors of Big Data. *International Journal of Communication*, 8, 1690-1709.
- Richards, N.M. & Hartzog, W. (2020). A Relational Turn for Data Protection? *4 European Data Protection Law Review*, 1, 1-6.
- Rickert, David, Hines, Walter, McKenzie, Stuart (1933) Methodology for River-Quality Assessment with Application to the Willamette River Basin, Oregon. *U.S. Geological Survey Circular*, 702-717, M1-M55.
- Rider, K. & Wood, D.M. (2018). Condemned to connection? Network communitarianism in Mark Zuckerberg's "Facebook Manifesto". *New Media & Society*, 21(3), 639-654.
- Rosenberg, D. (2013). Data before the fact. In L. Gitelman (Ed.), "Raw data" is an oxymoron (pp. 15-40). MIT Press
- Rossini, F. D. (1967). Historical background of data compiling activities. *Journal of Chemical Documentation*, 7(1), 2-6.
- Ruberg, B., & Ruelos, S. (2020). Data for queer lives: How LGBTQ gender and sexuality identities challenge norms of demographics. *Big Data & Society*, 1-12.
- Sarigol, E., Garcia, D. & Schweitzer, F. (2014). Privacy as a collective phenomenon. *ACM COSN '14*, 95-106.
- Sedgwick, E. (2003). *Touching Feeling: Affect, Pedagogy, Performativity*. Duke University Press.
- Spade, D. (2015). *Normal life: Administrative violence, critical trans politics, and the limits of law*. Duke University Press.
- Spivak, G.C. (1988). Can the Subaltern Speak? In C. Nelson and L. Grossberg (Eds.), *Marxism and the Interpretation of Culture* (pp. 24-28). Macmillan.
- Stark, L. & Hoffmann, A.L. (2019). Data Is the New What? Popular Metaphors & Professional Ethics in Emerging Data Culture. *Journal of Cultural Analytics*, 4549.
- Strycharz, J., Ausloos, J., & Helberger, N. (2020). Data Protection or Data Frustration? Individual Perceptions and Attitudes Towards the GDPR. *European Data Protection Law Review*, 6(3), 407-421.

Szulc, L. (2020). Digital Gender Disidentifications: Beyond the Subversion Versus Hegemony Dichotomy and Toward Everyday Gender Practices. *International Journal of Communication*, 14(2020), 5436-5454

Thalen, M. (2019). Facebook lawyer says users 'have no expectation of privacy'. Daily Dot. <https://www.dailydot.com/debug/facebook-lawyer-no-expectation-of-privacy/>.

Tisné, M. (2021). Collective data rights can stop big tech from obliterating privacy. *MIT Technology Review*. <https://www.technologyreview.com/2021/05/25/1025297/collective-data-rights-big-tech-privacy/>.

Toonders, J. (2014). Data is the New Oil of the Digital Economy. *Wired*. <https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/>.

Tutt, A. (2017). An FDA for algorithms. *Administrative Law Review*, 69, 83-124.

Wood, D.M. (2017). Urban Surveillance After the End of Globalization. In J. Short (Ed.), *A Research Agenda for Cities* (pp. 38-51). Edward Elgar.

Woodie, A. (2021). Drowning In a Data Lake? Gartner Analyst Offers a Life Preserver. *Datanami*. <https://www.datanami.com/2021/05/07/drowning-in-a-data-lake-gartner-analyst-offers-a-life-preserver/>.

[Funding]

No funding to declare.

[Declaration of novelty and no competing interests]

By submitting this manuscript I declare that this manuscript and its essential content has not been published elsewhere or that it is considered for publication in another outlet.

No competing interests exist that have influenced or can be perceived to have influenced the text.